

Output analysis for approximated stochastic programs

Jitka Dupačová (dupacova@karlin.mff.cuni.cz)

*Department of Probability and Mathematical Statistics, Charles University,
Sokolovská 83, CZ-186 75 Prague 8, Czech Republic*

Abstract

Because of incomplete information and also for the sake of numerical tractability one mostly solves an approximated stochastic program instead of the underlying "true" decision problem. However, without an additional analysis, the obtained output (the optimal value and optimal solutions of the approximated stochastic program) should not be used to replace the sought solution of the "true" problem. Methods of output analysis have to be tailored to the structure of the problem and they should also reflect the source, character and precision of the input data. The scope of various approaches based on results of asymptotic and robust statistics, of the moment problem and on general results of parametric programming will be discussed from the point of view of their applicability and possible extensions.

Keywords: Stochastic programming, approximated distributions, stability, sensitivity, robustness

1 Introduction

45 years ago, stochastic programming has initiated to reflect the observation that for many linear programs, to be solved, the values of the presumably known coefficients were not available. The deterministic view has been replaced by a stochastic one assuming that these coefficients or parameters are random and their probability distribution is *known* and independent of the decision variables.

The subsequent decades of theoretical analysis and software development for various types of stochastic programming models were influenced by developments in optimization, probability and statistics and in computer technologies, with the progress recorded step by step (for instance in [6, 22, 27, 59, 60]) and new tasks and challenges

have been opened, see e.g. [61]. However, from the very beginning there have been doubts if the basic assumption of a known probability distribution is a realistic one.

Let us consider a frequent framework for stochastic programs

$$\min F(\mathbf{x}, P) := E_P f_0(\mathbf{x}, \omega) \text{ on a set } \mathcal{X} \quad (1)$$

where P is a fixed probability distribution on (Ω, \mathcal{B}) , $\Omega \subset R^m$ which does not depend on \mathbf{x} , E_P is the corresponding expectation operator, $\mathcal{X} \subset R^n$ is a nonempty closed set which does not depend on P and the random objective f_0 is a rather complicated function $f_0(\mathbf{x}, \omega)$, $f_0 : \mathcal{X} \times \Omega \rightarrow R^1$. For instance, the values $f_0(\mathbf{x}, \omega)$ in two-stage stochastic programs are obtained as optimal values of certain second-stage mathematical programs.

We refer to the objective function in (1) as the *expectation functional* and to (1) as the *expectation-type stochastic program*; its objective function $F(\mathbf{x}, P)$ is linear in P . However, other choices of the functional $F(\mathbf{x}, P)$, for instance the convex-concave objective functions used in robust optimization [40] or the quantile function related with the VaR (value at risk) criterion, are also of interest.

To deal with incomplete information about P one can try to incorporate the available level of the (incomplete) information about the probability distribution P into the decision model. The first idea was to use the *minimax decision rule*. It means that instead of a specified probability distribution P a whole class \mathcal{P} of probability distributions is considered and the minimax (or the worst case) solution is a minimizer of

$$\sup_{P \in \mathcal{P}} E_P f_0(\mathbf{x}, \omega)$$

on the set \mathcal{X} . The definition of \mathcal{P} should reflect the available information mostly described by known values of moments. With the expectation functional in (1), this type of \mathcal{P} allows to exploit results of moment problems of mathematical statistics. However, except for rather special forms of $f_0(\mathbf{x}, \bullet)$ and for special choices of \mathcal{P} , the results are hard to apply for *decision making*. The reason is that except for these very special cases, the extremal probability distributions depend on the decision variables. For fixed $\mathbf{x} \in \mathcal{X}$ the results may be exploited in the context of the worst case analysis and approximation techniques. See Section 2.4 for discussions and references.

Another approach, connected also with the necessity to solve approximate stochastic programs instead of theoretical ones (based perhaps on a known probability distribution), is to analyze the output of the approximate stochastic program with the goal to draw inference about the output of the theoretical, true problem. Here, the probability distribution P plays a role of an abstract parameter which is estimated or approximated by another probability distribution \hat{P} obtained by parametric or nonparametric methods, by sampling, discretization and simulation techniques, or by accepting expert's view. The optimization problem (1) is solved with \hat{P} at the place of P and the question is how the obtained optimal value $\varphi(\hat{P})$ and the set of optimal solutions $\mathcal{X}^*(\hat{P})$ (or ε -optimal solutions) relate to the optimal value $\varphi(P)$ and

to the set of optimal solutions $\mathcal{X}^*(P)$ (or ε -optimal solutions) for the true problem (1). Depending on the nature of the approximation and also on structural properties of problem (1), various methods of asymptotic and robust statistics, of parametric programming and also simulation techniques can be used; in general, it is much easier to estimate the precision of the obtained optimal value than that of optimal solutions. The relevant approaches will be sum up under the name *output analysis*. To be correct and efficient, methods of output analysis have to be tailored to the nature of the solved real-life problem, to the structure of the chosen stochastic programming model and they should also reflect the source, character and precision of the input data.

In the sequel, we shall discuss basic asymptotic results (Section 2.1), we shall mention results on qualitative and quantitative stability (Section 2.2) and we shall deal with contamination technique in Section 2.3. We shall focus on assumptions used to get the theoretical results and we shall specify the classes of stochastic programs to which these approaches can be applied. We shall indicate several extensions to stochastic programs with $F(\mathbf{x}, P)$ nonlinear in P or nonconvex in \mathbf{x} . On the other hand, neither simulation methods nor various bounding techniques will be discussed here even when they can also be exploited for purposes of output analysis.

2 Methods of output analysis

2.1 Asymptotic inference

is based on the assumption that the true probability distribution P in (1) can be well approximated by an infinite sequence of probability distributions based on an increasing level of information about P . This can be modeled in the following way: Consider a *sample space* (Z, \mathcal{F}, μ) with an increasing sequence of σ -fields $(\mathcal{F}^\nu)_{\nu=1}^\infty$ contained in \mathcal{F} . A sample ζ leads to a sequence of \mathcal{F}^ν -measurable probability distributions $\{P^\nu(\bullet, \zeta), \nu = 1, 2, \dots\}$ on (Ω, \mathcal{B}) that are based on the information collected up to ν . The optimal value φ and the optimal solutions \mathbf{x}^* of the approximate stochastic program

$$\min_{\mathbf{x} \in \mathcal{X}} F(\mathbf{x}, P^\nu) = \min_{\mathbf{x} \in \mathcal{X}} E_{P^\nu} f_0(\mathbf{x}, \omega) \quad (2)$$

based on $P^\nu(\bullet, \zeta)$ depend on the used sample path ζ and the *presented results hold true for almost all sample paths* ζ , i.e., μ -a.s. The probability distributions P^ν will be called the *wide-sense empirical* probability distributions and the same designation will be used also for the approximate stochastic programs (2), their optimal values and optimal solutions. This will help to distinguish among general asymptotic results and those valid for the empirical probability distributions. In the latter case, the sample path $\zeta = \{\omega^1, \omega^2, \dots\}$ is obtained by simple random sampling from (Ω, \mathcal{B}, P) , $\mu = P^\infty$

and the *empirical stochastic program* is

$$\min_{\mathbf{x} \in \mathcal{X}} F(\mathbf{x}, P^\nu) = \min_{\mathbf{x} \in \mathcal{X}} \frac{1}{\nu} \sum_{j=1}^{\nu} f_0(\mathbf{x}, \omega^j). \quad (3)$$

2.1.1 Consistency results

Under assumption that $P^\nu \rightarrow P$ weakly and that $f_0(\mathbf{x}, \bullet)$ is a continuous bounded function of ω for every $\mathbf{x} \in \mathcal{X}$, the pointwise convergence of the expected value objectives $F(\mathbf{x}, P^\nu) \rightarrow F(\mathbf{x}, P) \forall \mathbf{x} \in \mathcal{X}$ follows directly from the definition of weak convergence. If \mathcal{X} is *compact* and the convergence of the expectations is *uniform* on \mathcal{X} we get immediately (μ -a.s.) convergence of the optimal values

$$\varphi(P^\nu) \rightarrow \varphi(P).$$

If, moreover, \mathcal{X} is *convex* and $f_0(\bullet, \omega)$ is *strictly convex* on \mathcal{X} it is easy to get convergence of the (unique) optimal solutions $\mathbf{x}^*(P^\nu)$ of $\min_{\mathbf{x} \in \mathcal{X}} F(\mathbf{x}, P^\nu)$ to the unique optimal solution $\mathbf{x}^*(P)$ of the initial problem (1) and some rates of the convergence. This approach was used in early papers on consistency, e.g. [32]. Notice that merely the pointwise convergence of the empirical expectations does not imply consistency of the optimal values.

Another way is via epi-convergence of lower semicontinuous (lsc) functions. The main step is to prove that the approximate objective functions $F(\mathbf{x}, P^\nu)$ epi-converge to the true objective function in (1), which in turn implies convergence results for the optimal values and of the sets of optimal and ε -optimal solutions, cf. [48]. Let us recall the assumptions.

Basic assumptions:

- a. $\mathcal{X} \subset R^n$ is a nonempty closed set,
- b. $f_0(\mathbf{x}, \omega)$ is a random lower semicontinuous function; this means that f_0 is jointly measurable and $f_0(\bullet, \omega)$ is lower semicontinuous for all $\omega \in \Omega$,
- c. $P^\nu \rightarrow P$ weakly.

To get *epi-convergence* of the expectations $F(\mathbf{x}, P^\nu) \rightarrow F(\mathbf{x}, P)$, additional assumptions concerning the convergence of $P^\nu \rightarrow P$ and the properties of f_0 are needed. The assumptions mimic to a certain extent those of the classical consistency result:

- d. *continuity* of $f_0(\mathbf{x}, \bullet)$ on Ω ,
- e. *uniform convergence* (asymptotic negligibility, tightness) of probability distributions P, P^ν with respect to the functions $f_0(\mathbf{x}, \bullet) \forall \mathbf{x} \in \mathcal{X}$; this replaces the former assumption of bounded integrals $f_0(\mathbf{x}, \bullet) \forall \mathbf{x}$.
- f. *local (lower) Lipschitz property* of $f_0(\bullet, \omega)$ for all $\omega \in \Omega$; in case of $f_0(\bullet, \omega)$ convex for all $\omega \in \Omega$, this assumption is not needed.

If these assumptions are fulfilled, we get

Proposition 1. (cf. Theorems 3.7, 3.8 of [20]): *Under assumptions a-f, $F(\mathbf{x}, P)$ is μ -a.s. both epi-limit and pointwise limit of $F(\mathbf{x}, P^\nu)$ for $\nu \rightarrow \infty$.*

Epi-convergence of the objective functions, see [48], implies the consistency result:

Proposition 2. (cf. Theorem 3.9 of [20]): *Under assumptions **a–f** we have that μ -a.s.*

$$\limsup_{\nu \rightarrow \infty} \varphi(P^\nu) \leq \varphi(P)$$

and any cluster point $\hat{\mathbf{x}}$ of any sequence $\{\mathbf{x}^\nu, \nu = 1, 2, \dots\}$ with $\mathbf{x}^\nu \in \mathcal{X}^(P^\nu)$ belongs to $\mathcal{X}^*(P)$.*

In particular, if there is a compact set $\mathcal{D} \subset R^n$ such that μ -a.s., $\mathcal{X}^(P^\nu) \cap \mathcal{D} \neq \emptyset$ for $\nu = 1, 2, \dots$ and $\{\mathbf{x}^*\} \in \mathcal{X}^*(P) \cap \mathcal{D}$ then there exist measurable selection \mathbf{x}^ν of $\mathcal{X}^*(P^\nu)$ such that $\mathbf{x}^* = \lim_{\nu \rightarrow \infty} \mathbf{x}^\nu$ for μ -almost all ζ and also $\varphi(P) = \lim_{\nu \rightarrow \infty} \varphi(P^\nu)$ μ -a.s.*

Application 1. Consistency result for a discrete true distribution.

Let $\omega^1, \dots, \omega^N$ be the atoms of P and $\pi_j > 0$, $j = 1, \dots, N$, $\sum_{j=1}^N \pi_j = 1$ their probabilities, let \mathcal{X} be a nonempty bounded convex polyhedron and $f_0(\mathbf{x}, \omega)$ a continuous function of ω on $\text{conv}\{\omega^1, \dots, \omega^N\}$ and a piece-wise linear convex function of \mathbf{x} on \mathcal{X} , i.e., the type of the random objective function which is common for two-stage stochastic linear programs. This implies that $F(\mathbf{x}, P) := \sum_{j=1}^N \pi_j f_0(\mathbf{x}, \omega^j)$ is also a piece-wise linear convex function, hence, there exists a finite number of bounded nonoverlapping convex polyhedra \mathcal{X}^k , $k = 1, \dots, K$, such that $\mathcal{X} = \cup_{k=1}^K \mathcal{X}^k$ and $F(\mathbf{x}, P)$ is linear on $\mathcal{X}^k \forall k$. Then the set of optimal solutions $\mathcal{X}^*(P)$ evidently intersects the set $\overline{\mathcal{X}}(P)$ of all extremal points of \mathcal{X}^k , $k = 1, \dots, K$.

The true distribution P is estimated by empirical distributions P^ν based on finite samples of sizes ν from P , hence, carried by subsets of $\{\omega^1, \dots, \omega^N\}$. The empirical objective functions $F(\mathbf{x}, P^\nu)$ are also convex, piece-wise linear and the sets of the related extremal points $\overline{\mathcal{X}}(P^\nu) \subset \overline{\mathcal{X}}(P)$. This means that the assumptions of Proposition 2 are fulfilled with the compact set $\mathcal{D} = \overline{\mathcal{X}}(P)$. Consequently, *with probability one, any cluster point of any sequence of points $\mathbf{x}^\nu \in \mathcal{X}^*(P^\nu) \cap \mathcal{D}$ is an optimal solution of the true problem.*

Assume in addition that there is a unique optimal solution $\mathbf{x}^*(P)$ of the true problem

$$\min_{\mathbf{x} \in \mathcal{X}} \sum_{j=1}^N \pi_j f_0(\mathbf{x}, \omega^j)$$

In this case there is a measurable selection $\mathbf{x}^*(P^\nu)$ from $\mathcal{X}^*(P^\nu) \cap \mathcal{D}$ such that $\lim_{\nu \rightarrow \infty} \mathbf{x}^*(P^\nu) = \mathbf{x}^*(P)$ μ -a.s. Due to the special form of the objective functions and of the sets $\mathcal{X}^*(P^\nu) \cap \mathcal{D}$, this is equivalent to

$$\mathbf{x}^*(P^\nu) \equiv \mathbf{x}^*(P) \quad \mu - \text{a.s. for } \nu \text{ large enough.}$$

This means that *for ν large enough the empirical problem provides with probability 1 the exact optimal solution of the true problem.* This result can be extended to the case of multiple optimal solutions and for $\nu \rightarrow \infty$ the probability approaches 1 exponentially fast, cf. [55].

Comments

- For *convex* function $f_0(\bullet, \omega)$, convex set \mathcal{X} and for empirical probability distributions, P^ν epi-convergence of $F(\mathbf{x}, P^\nu)$ to $F(\mathbf{x}, P)$ follows from the strong law of large numbers for sums of random closed sets and the consistency result can be extended from R^n to reflexive Banach spaces [36].
- Another type of extension to infinite dimensional spaces can be found in [39].
- There are various further results focused on wide-sense empirical stochastic programs under various assumptions about the approximate probability distributions P^ν ; see for instance [37].
- The approach based on epi-convergence can be applied to problems where both the integrand and the probability distribution are approximated, see e.g. [46].
- An important generalization is to *discontinuous integrands* $f_0(\mathbf{x}, \bullet)$. In such cases, uniform integrability is not sufficient for semicontinuity of integrals $F(\mathbf{x}, P^\nu)$. A suitable additional condition is that the probability of the set of discontinuity points of the integrand $f_0(\mathbf{x}, \bullet)$ for the true problem is zero; see [1]. A motivation comes for instance from approximated integer stochastic programs, cf. [52].
- Possible extensions to objective functions $F(\mathbf{x}, P)$ which are nonlinear in P are in progress.
- The consistency result can be exploited also in case that some of constraints are in the form of expectations, e.g., Theorem 6.3 in [60]. This in turn helps to get consistency for stochastic programs with *quantile-type objective functions* in which the true probability distributions are approximated by the empirical ones.
- Theoretically, the above asymptotic results apply also to the *multistage* stochastic programs. However, the assumption of an infinitely increasing sample size means that at every node of the scenario tree, the number of branches grows to infinity and the sample based problems become very quickly intractable.

2.1.2 Asymptotic distributions and rates of convergence

are derived under assumption that the corresponding consistency results hold true. An instance of a *large deviations* result of [55] was mentioned in the context of Application 1, further results can be found in [31]. We shall discuss here briefly results on asymptotic distributions.

Asymptotic normality of the optimal value $\varphi(P^\nu)$ of empirical stochastic program (3) can be proved under relatively weak assumptions, e.g., for compact $\mathcal{X} \neq \emptyset$, unique true optimal solution $\mathbf{x}^*(P)$ and for $f_0(\bullet, \omega)$ Lipschitz continuous $\forall \omega$, with finite expectation $E_P f_0^2(\mathbf{x}^*(P), \omega)$, see Theorem 3.3 of [54]. This allows to construct approximate confidence intervals and tests concerning the true optimal value.

Asymptotic distribution of optimal solutions. The first ideas concerning asymptotic distribution of the empirical optimal solutions were delineated in [58] for unconstrained stochastic programs. It turns out, however, that in the presence of constraints asymptotic normality of empirical optimal solutions $\mathbf{x}^*(P^\nu)$ cannot be expected even

when all solution sets $\mathcal{X}^*(P)$ and $\mathcal{X}^*(P^\nu) \forall \nu$ are singletons. It is possible to prove that under reasonable assumptions, the asymptotic distribution of optimal solutions $\mathbf{x}^*(P^\nu)$ is *conically normal* being projection of a normal distribution on a convex cone, whereas asymptotic normality can appear only when the problem reduces on a neighborhood of the true optimal solution $\mathbf{x}^*(P)$ to an unconstrained one. The general tool is the generalized δ -method, introduced in [34]. Similarly as the classical δ -method, it requires certain differentiability property of the mapping in question, i.e., of the optimal solution map \mathbf{x}^* at P , and a suitable version of the Central Limit Theorem. Limiting conditions under which asymptotic normality of wide-sense empirical optimal solutions holds true can be found in [20].

We shall mention below only some of ideas, formulated here under relatively strong assumptions; see [34, 35] for a full derivation of the related asymptotic results.

Assume for simplicity that $\mathcal{X} \subset R^n$ is convex, compact, polyhedral, $\text{int}\mathcal{X} \neq \emptyset$ and $f_0(\bullet, \omega)$ is a convex function of \mathbf{x} . Then $\mathbf{x}^*(P)$ is an optimal solution of (1) iff

$$\mathbf{0} \in \partial_{\mathbf{x}} F(\mathbf{x}^*(P), P) + N_{\mathcal{X}}(\mathbf{x}^*(P)), \quad (4)$$

i.e., iff there is $\mathbf{g}_0(P) \in \partial_{\mathbf{x}} F(\mathbf{x}^*(P), P)$ and $\mathbf{g}_{\mathcal{X}}(P)$ belonging to the normal cone $N_{\mathcal{X}}(\mathbf{x}^*(P))$ such that

$$\mathbf{0} = \mathbf{g}_0(P) + \mathbf{g}_{\mathcal{X}}(P).$$

The relevant results follow partly by analysis of the first order optimality conditions (4). If $F(\bullet, P)$ is differentiable, (4) implies that

$$-\nabla_{\mathbf{x}} F(\mathbf{x}^*(P), P) \in N_{\mathcal{X}}(\mathbf{x}^*(P))$$

and the necessary condition for differentiability of \mathbf{x}^* at P (in the sense of an *affine* approximation of its graph at P) reads

$$-\nabla_{\mathbf{x}} F(\mathbf{x}^*(P), P) \in \text{int}N_{\mathcal{X}}(\mathbf{x}^*(P)), \quad (5)$$

cf. [44]. Condition (5) allows to prove asymptotic normality of consistent wide-sense empirical estimates $\mathbf{x}^*(P^\nu)$. The reason is clear: If (5) holds true, problem (1) reduces locally to an unconstrained one. Also, (5) implies, see [12], that the limiting assumption for asymptotic normality as formulated in 4.7 (ii) from [20]

$$\sqrt{\nu}(\mathbf{g}_{\mathcal{X}}(P) - \mathbf{g}_{\mathcal{X}}(P^\nu)) \rightarrow 0$$

in probability is fulfilled.

In the sequel we shall concentrate on the empirical problems. We shall assume that the integrands $f_0(\bullet, \omega)$ are continuously differentiable, that the matrix $\mathbf{G}(\mathbf{x}^*(P)) := E_P \nabla_{xx}^2 f_0(\mathbf{x}^*(P), \omega)$ exists and is positively definite and that the Central Limit Theorem for continuous functions holds true for $\nabla_{\mathbf{x}} F(\mathbf{x}, P^\nu)$. Then there exists a unique optimal solution $\mathbf{x}^*(P)$ of (4) and

$$\sqrt{\nu}(\nabla_{\mathbf{x}} F(\mathbf{x}, P^\nu) - \nabla_{\mathbf{x}} F(\mathbf{x}, P)) \sim \mathcal{N}(\mathbf{0}, \mathbf{V}(\mathbf{x})) \quad (6)$$

with variance matrix $\mathbf{V}(\mathbf{x}) = \text{var}\{\nabla_x f_0(x, \omega)\}$.

The final statement, see [34, 35], is that the (unique) optimal solution of the following randomly perturbed quadratic program

$$\min_{\mathbf{u} \in T_0(\mathbf{x}^*(P))} \mathbf{u}^\top \mathbf{w}(\mathbf{x}^*(P)) + \frac{1}{2} \mathbf{u}^\top \mathbf{G}(\mathbf{x}^*(P)) \mathbf{u} \quad (7)$$

with normally $\mathcal{N}(\mathbf{0}, \mathbf{V}(\mathbf{x}))$ distributed $\mathbf{w}(\mathbf{x})$ and with $T_0(\mathbf{x}^*(P))$ the critical cone at $\mathbf{x}^*(P)$ is a generalized directional derivative of \mathbf{x}^* at P to be used in the δ -method. As a result, *the asymptotic distribution of $\sqrt{\nu}(\mathbf{x}^*(P^\nu) - \mathbf{x}^*(P))$ is that of the projection with respect to $\mathbf{G}(\mathbf{x}^*(P))$ of the asymptotically normal vector*

$$\boldsymbol{\varrho}^*(P) := \mathbf{G}(\mathbf{x}^*(P))^{-1} \mathbf{w}(\mathbf{x}^*(P)), \quad \mathbf{w}(\mathbf{x}^*(P)) \sim \mathcal{N}(\mathbf{0}, \mathbf{V}(\mathbf{x}^*(P)))$$

on the critical cone $T_0(\mathbf{x}^*(P))$.

Together with stability theory for quadratic programs, see for instance Chapter 5 of [3], this result provides a clear characterization of the conically normal distribution. The asymptotic distribution is normal iff $\boldsymbol{\varrho}^*(P)$ is an interior point of one of (finitely many, polyhedral) stability sets. This means in turn that asymptotic normality of empirical solutions will appear quite often.

Comments

- The central limit theorem for $\nabla_x F(\mathbf{x}, P^\nu)$ is obtained, e.g., for $f_0(\bullet, \omega)$ convex $C^{1,1}$ -functions for all ω , with square integrable Lipschitz constants, with a finite nonsingular variance matrix $\mathbf{V}(\mathbf{x}^*(P)) = \text{var}[\nabla_x f_0(\mathbf{x}^*(P), \omega)]$, a finite expectation $E\|\nabla_x f_0(\mathbf{x}^*(P), \omega)\|^2$ and for empirical probability distributions P^ν . The assumption of empirical probability distributions can be relaxed in various ways; an example is exploitation of a Central Limit Theorem for time series in [56].

- For nonnormal asymptotic distributions of the empirical expectation functionals and for rates of convergence different from $(\sqrt{\nu})^{-1}$ see [42].

- Differentiability assumptions concerning $f_0(\mathbf{x}, \omega)$ or $F(\mathbf{x}, P)$ restrict applicability of these results. To an extent, these assumptions can be relaxed; see for instance [35].

- In case of explicit constraints one exploits mostly optimality conditions based on the classical Lagrangian approach; see, e.g., [35] or Chapter 6 of [50]. For this type of problems, conically normal asymptotic distribution of optimal solutions of approximate problems can be obtained for certain types of wide-sense empirical programs, cf. [57]. The required key property is again validity of a version of the Central Limit Theorem for (generalized) gradients of the empirical expectation functionals.

2.1.3 Asymptotic results for a parametric family

A simplification is possible whenever the general stability properties with respect to the probability distribution can be reduced to a finite dimensional parameter case. An example are probability distributions of a given parametric form and the desired

results concern differences between the optimal values $\varphi(\boldsymbol{\theta}_0)$ and $\varphi(\boldsymbol{\theta}_\nu)$ and between the solution sets obtained for the true parameter value $\boldsymbol{\theta}_0$ and for its estimate $\boldsymbol{\theta}_\nu$, respectively. For sufficiently smooth optimal value function φ and for unique optimal solutions, the statistical properties of $\varphi(\boldsymbol{\theta}_\nu)$ and of $\mathbf{x}^*(\boldsymbol{\theta}_\nu)$ follow from the statistical properties of the estimates $\boldsymbol{\theta}_\nu$ by application of results concerning transformed random sequences [53]. This was elaborated in [8]; see also [10] and references ib.

Concerning asymptotic properties of the unique optimal solutions $\mathbf{x}^*(\boldsymbol{\theta}_\nu)$, differentiability at $\boldsymbol{\theta}_0$ holds true under a condition parallel to (5). Otherwise, a quadratic program similar to (7) provides directional derivatives of \mathbf{x}^* at $\boldsymbol{\theta}_0$, in the direction of $\sqrt{\nu}(\boldsymbol{\theta}_\nu - \boldsymbol{\theta}_0)$. For the purposes of optimization, the set \mathcal{X} can be approximated for \mathbf{x} near to $\mathbf{x}^*(\boldsymbol{\theta}_0)$ by the critical cone $T_0 := T_0(\mathbf{x}^*(\boldsymbol{\theta}_0))$. Under additional assumptions of differentiability of the gradients $\nabla_x F(\mathbf{x}, \boldsymbol{\theta})$ with respect to both \mathbf{x} and $\boldsymbol{\theta}$, with a positively definite matrix $\mathbf{G}(\boldsymbol{\theta}_0) := \nabla_{x\theta}^2 F(\mathbf{x}^*(\boldsymbol{\theta}_0), \boldsymbol{\theta}_0)$, the generalized equation

$$\mathbf{0} \in \nabla_x F(\mathbf{x}, \boldsymbol{\theta}) + N_{\mathcal{X}}(\mathbf{x})$$

can be approximated for \mathbf{x} near to $\mathbf{x}^*(\boldsymbol{\theta}_0)$ and $\boldsymbol{\theta}$ near to $\boldsymbol{\theta}_0$ by the *linearized* generalized equation

$$\mathbf{0} \in \nabla_x F(\mathbf{x}^*(\boldsymbol{\theta}_0), \boldsymbol{\theta}_0) + \nabla_{x\theta}^2 F(\mathbf{x}^*(\boldsymbol{\theta}_0), \boldsymbol{\theta}_0)(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + \mathbf{G}(\boldsymbol{\theta}_0)(\mathbf{x} - \mathbf{x}^*(\boldsymbol{\theta}_0)) + N_{T_0}(\mathbf{x}).$$

Hence, the difference $\mathbf{x}^*(\boldsymbol{\theta}) - \mathbf{x}^*(\boldsymbol{\theta}_0)$ is approximated on a neighborhood of $\mathbf{0}$ by the unique solution of the quadratic program

$$\min_{\mathbf{u} \in T_0(\mathbf{x}^*(\boldsymbol{\theta}_0))} \mathbf{u}^\top [\nabla_x F(\mathbf{x}^*(\boldsymbol{\theta}_0), \boldsymbol{\theta}_0) + \nabla_{x\theta}^2 F(\mathbf{x}^*(\boldsymbol{\theta}_0), \boldsymbol{\theta}_0)(\boldsymbol{\theta} - \boldsymbol{\theta}_0)] + \frac{1}{2} \mathbf{u}^\top \mathbf{G}(\boldsymbol{\theta}_0) \mathbf{u}. \quad (8)$$

According to [44], Theorem 3.5, this optimal solution equals the directional derivative $D\mathbf{x}^*(\boldsymbol{\theta}_0; \boldsymbol{\theta} - \boldsymbol{\theta}_0)$ of the optimal solution $\mathbf{x}^*(\boldsymbol{\theta}_0)$ in the direction of $\boldsymbol{\theta} - \boldsymbol{\theta}_0$; on the critical cone, the first linear term in the objective function of (8) vanishes.

Consider now $\boldsymbol{\theta}_\nu, \nu = 1, 2, \dots$ a sequence of consistent and asymptotically normal estimates of $\boldsymbol{\theta}_0$, i.e., $\sqrt{\nu}(\boldsymbol{\theta}_\nu - \boldsymbol{\theta}_0) \sim \mathcal{N}(\mathbf{0}, \mathbf{V})$, denote \mathbf{w} the corresponding normally distributed random vector and rewrite the quadratic program (8) with $\boldsymbol{\theta} = \boldsymbol{\theta}_0 + \sqrt{\nu}(\boldsymbol{\theta}_\nu - \boldsymbol{\theta}_0)$. We get randomly perturbed quadratic program

$$\min_{\mathbf{u} \in T_0(\mathbf{x}^*(\boldsymbol{\theta}_0))} \mathbf{u}^\top [\nabla_x F(\mathbf{x}^*(\boldsymbol{\theta}_0), \boldsymbol{\theta}_0) + \sqrt{\nu} \nabla_{x\theta}^2 F(\mathbf{x}^*(\boldsymbol{\theta}_0), \boldsymbol{\theta}_0)(\boldsymbol{\theta}_\nu - \boldsymbol{\theta}_0)] + \frac{1}{2} \mathbf{u}^\top \mathbf{G}(\boldsymbol{\theta}_0) \mathbf{u}. \quad (9)$$

The asymptotic distribution of the (unique) optimal solution of (9) is asymptotically equivalent to the asymptotic distribution of $\sqrt{\nu}(\mathbf{x}^*(\boldsymbol{\theta}_\nu) - \mathbf{x}^*(\boldsymbol{\theta}_0))$ and to the probability distribution of the optimal solution $\mathbf{u}(\mathbf{w}; \boldsymbol{\theta}_0)$ of the random quadratic program

$$\min_{\mathbf{u} \in T_0(\mathbf{x}^*(\boldsymbol{\theta}_0))} \mathbf{u}^\top [\nabla_x F(\mathbf{x}^*(\boldsymbol{\theta}_0), \boldsymbol{\theta}_0) + \nabla_{x\theta}^2 F(\mathbf{x}^*(\boldsymbol{\theta}_0), \boldsymbol{\theta}_0) \mathbf{w}] + \frac{1}{2} \mathbf{u}^\top \mathbf{G}(\boldsymbol{\theta}_0) \mathbf{u}. \quad (10)$$

2.1.4 The approximated random quadratic program

We have learned that asymptotic distribution of optimal solutions of empirical stochastic programs is that of optimal solutions of random quadratic programs of the form (7) or (10). However, an exploitation of this result in output analysis is not straightforward. The problem is that the coefficients in the objective function and also the critical cones T_0 and the variance matrices \mathbf{V} depend on the unknown optimal solutions $\mathbf{x}^*(P)$ or $\mathbf{x}^*(\boldsymbol{\theta}_0)$ of the corresponding true stochastic programs.

The first idea is to replace the critical cone in (7) or (10) by the convex polyhedral set \mathcal{X} shifted by the true solution $\mathbf{x}^*(P)$ or $\mathbf{x}^*(\boldsymbol{\theta}_0)$; on a sufficiently small neighborhood of zero, the optimal solutions of the related pairs of quadratic programs are identical; see discussion in [44] and [11].

The next suggestion of [11] is to replace the random quadratic program by its empirical counterpart. In the parametric case, the two suggested approximations of (10) lead to random quadratic program

$$\min_{\mathbf{u} \in \mathcal{X} - \{\mathbf{x}^*(\boldsymbol{\theta}_\nu)\}} \mathbf{u}^\top [\nabla_x F(\mathbf{x}^*(\boldsymbol{\theta}_\nu), \boldsymbol{\theta}_\nu) + \nabla_{x\theta}^2 F(\mathbf{x}^*(\boldsymbol{\theta}_\nu), \boldsymbol{\theta}_\nu) \mathbf{w}_\nu] + \frac{1}{2} \mathbf{u}^\top \mathbf{G}(\boldsymbol{\theta}_\nu) \mathbf{u} \quad (11)$$

with normally $\mathcal{N}(\mathbf{0}, \mathbf{V}_\nu)$ distributed perturbation \mathbf{w}_ν obtained by replacing the variance matrix \mathbf{V} by its empirical counterpart \mathbf{V}_ν . If the derivatives of $\nabla_x F(\mathbf{x}, \boldsymbol{\theta})$ with respect to \mathbf{x} and $\boldsymbol{\theta}$ are continuous, if the matrices \mathbf{V}_ν and $\mathbf{G}(\boldsymbol{\theta}_\nu) := \nabla_{xx}^2 F(\mathbf{x}^*(\boldsymbol{\theta}_\nu), \boldsymbol{\theta}_\nu)$ are positive definite for all ν , the (unique) optimal solutions $\mathbf{u}(\mathbf{w}_\nu; \boldsymbol{\theta}_\nu)$ of (11) are continuous in all coefficients of the objective function and in the shift of the set \mathcal{X} , cf. [3], Section 5.3. As all coefficients converge in distribution to the coefficients of the initial random quadratic program, the optimal solutions $\mathbf{u}(\mathbf{w}_\nu; \boldsymbol{\theta}_\nu)$ of (11) are asymptotically equivalent to the optimal solutions of quadratic program (10). It means that *the asymptotic distribution of $\sqrt{\nu}(\mathbf{x}^*(\boldsymbol{\theta}_\nu) - \mathbf{x}^*(\boldsymbol{\theta}_0))$ equals the asymptotic distribution of the optimal solution $\mathbf{u}(\mathbf{w}_\nu; \boldsymbol{\theta}_\nu)$ of the approximated quadratic program (11)*; see [11] for a more detailed discussion.

In view of the local validity of the approximation of the critical cone by the shifted set \mathcal{X} (and vice versa) we might be interested to test if $\mathbf{x}^*(\boldsymbol{\theta}_\nu)$ is sufficiently close to the true solution $\mathbf{x}^*(\boldsymbol{\theta}_0)$. As we already know $\mathbf{u}(\mathbf{w}_\nu; \boldsymbol{\theta}_\nu)$ are $\mathbf{G}(\boldsymbol{\theta}_\nu)$ -projections of asymptotically normal vectors – *unconstrained* minimizers of (11) –

$$\boldsymbol{\varrho}_\nu := -\mathbf{G}^{-1}(\boldsymbol{\theta}_\nu) [\nabla_x F(\mathbf{x}^*(\boldsymbol{\theta}_\nu), \boldsymbol{\theta}_\nu) + \nabla_{x\theta}^2 F(\mathbf{x}^*(\boldsymbol{\theta}_\nu), \boldsymbol{\theta}_\nu) \mathbf{w}_\nu] \quad (12)$$

on the critical cone $T_0(\mathbf{x}^*(\boldsymbol{\theta}_\nu))$ or on the shifted set $\mathcal{X} - \{\mathbf{x}^*(\boldsymbol{\theta}_\nu)\}$. The two projections need not be identical. The suggested approximations are precise enough only on a sufficiently small neighborhood of the true solution point. The required size of such neighborhood depends also on the shape of the set \mathcal{X} around $\mathbf{x}^*(\boldsymbol{\theta}_0)$ and on its size when compared with the stochastic variability of $\boldsymbol{\varrho}_\nu$.

Variability of any n -dimensional random vector whose probability distribution is centered at \mathbf{m} and the variance matrix is \mathbf{C} can be described by the *concentration*

ellipsoid

$$\mathcal{E}(\mathbf{m}, \mathbf{C}; n+2) := \{\mathbf{z} \in R^n \mid (\mathbf{z} - \mathbf{m})^\top \mathbf{C}^{-1}(\mathbf{z} - \mathbf{m}) \leq n+2\} \quad (13)$$

constructed so that the uniform distribution on $\mathcal{E}(\mathbf{m}, \mathbf{C}; n+2)$ has the same expectation \mathbf{m} and the same variance matrix \mathbf{C} . The volume of the concentration ellipsoid (13) is $\frac{\pi^{n/2}(n+2)^{n/2}}{\Gamma(n/2+1)}(\det \mathbf{C})^{1/2}$ so that $\det \mathbf{C}$, called the *generalized variance*, characterizes the volatility of the random vector. In our case, see (12), the expectation of \mathbf{g}_ν is $\mathbf{m}_\nu = -\mathbf{G}^{-1}(\boldsymbol{\theta}_\nu) \nabla_x F(\mathbf{x}^*(\boldsymbol{\theta}_\nu), \boldsymbol{\theta}_\nu)$ and the variance matrix is

$$\mathbf{C}_\nu = \mathbf{G}^{-1}(\boldsymbol{\theta}_\nu) \nabla_{x,\theta}^2 F(\mathbf{x}^*(\boldsymbol{\theta}_\nu), \boldsymbol{\theta}_\nu) \mathbf{V}_\nu (\nabla_{x,\theta}^2 F(\mathbf{x}^*(\boldsymbol{\theta}_\nu), \boldsymbol{\theta}_\nu)^\top \mathbf{G}^{-1}(\boldsymbol{\theta}_\nu).$$

A possible test of asymptotic normality of optimal solutions and of feasibility of the approximation should consist of two main steps:

- (i) Solution of the unperturbed approximated quadratic program, i.e., projection of $\mathbf{p}_\nu := -\mathbf{G}^{-1}(\boldsymbol{\theta}_\nu) \nabla_x F(\mathbf{x}^*(\boldsymbol{\theta}_\nu), \boldsymbol{\theta}_\nu)$ on the set $\mathcal{X} - \{\mathbf{x}^*(\boldsymbol{\theta}_\nu)\}$. This step identifies, i.a., the face which contains optimal solutions of the quadratic program (11) for all sufficiently small perturbations of \mathbf{p}_ν provided that \mathbf{p}_ν is an interior point of a stability set.
- (ii) Projection of the concentration ellipsoid on the set $\mathcal{X} - \{\mathbf{x}^*(\boldsymbol{\theta}_\nu)\}$. If the projection is contained in the face identified in the step (i), both the hypothesis of asymptotic normality and the feasibility of the approximation are likely to hold true. If not, then the test should be repeated with an increased sample size.

Evidently, one could base the test also on *confidence* ellipsoids.

2.2 Qualitative and quantitative stability results

For various reasons, empirical estimates of the probability distribution P or of the true parameter $\boldsymbol{\theta}_0$ are not always available and, moreover, they need not provide the best approximation technique: They focus solely on the probability distribution, which is not the only ingredient of the stochastic programming models, they do not take into account any expert knowledge or foresight and for technical reasons, they cannot be based on very large samples. Moreover, the goal is to get a sensible approximation of the optimal solution and of the optimal value, not an approximation of the probability distribution. We shall look now into stability analysis of (1) with respect to the abstract parameter P .

2.2.1 Qualitative stability results

can be found in papers [28, 47]. They are based on general results of [45] and they have been specified to expectation-type stochastic programs (1). Certain assumption

of *joint* continuity of the objective function $F(\mathbf{x}, P)$ in \mathbf{x} and P is required whereas for stability of local minimizers, assumption of convexity of program (1) is not needed. One deals instead with complete local minimizing (CLM) sets:

For a given probability distribution $P = P_0$ the set \mathcal{M} is called a *complete local minimizing set* (CLM set) for (1) with respect to an open bounded set $\mathcal{G} \subset R^n$ if $\mathcal{M} \subset \mathcal{G}$ and \mathcal{M} is the set of all local minimizers of $F(\mathbf{x}, P_0)$ on the closure of \mathcal{G} . We denote

$$\varphi_{\mathcal{G}}(P_0) := \inf\{F(\mathbf{x}, P_0) | \mathbf{x} \in \text{cl}\mathcal{G}\}$$

the optimal value of (1) restricted to the set $\text{cl}\mathcal{G}$ and by $\mathcal{X}_{\mathcal{G}}^*(P_0)$ the corresponding set of optimal solutions. We assume that $\mathcal{X}_{\mathcal{G}}^*(P_0)$ is a CLM set for P_0 with respect to \mathcal{G} . Suitable continuity assumptions concerning $F(\mathbf{x}, P)$ imply that *the optimal value $\varphi_{\mathcal{G}}$ is finite, continuous at P_0 , the CLM set $\mathcal{X}_{\mathcal{G}}^*$ is Berge upper semicontinuous at P_0 and $\mathcal{X}_{\mathcal{G}}^*(P)$ is a nonempty compact CLM set with respect to \mathcal{G} for all P belonging to a neighborhood of P_0 .*

The required continuity properties of $F(\mathbf{x}, P)$ result, for instance, from joint continuity of f_0 on $\text{cl}\mathcal{G} \times R^m$ and uniform integrability of $f_0(\mathbf{x}, \omega)$, $\mathbf{x} \in \text{cl}\mathcal{G}$, with respect to the considered class of probability distributions. However, they can be obtained also under relaxed continuity properties of the integrands f_0 . An example is Proposition 4.1 of [51] which exploits structural properties of the two-stage stochastic integer programs in question and restricts the probability distribution of random coefficients to a specified class $\Delta_{p,K}$ for which the total absolute moments of order $p > 1$ are bounded by $K > 0$.

Comments

- Evidently, qualitative stability results for the optimal value function of empirical stochastic programs can be related to the corresponding consistency results.
- The general approach of [45] does not require linearity of $F(\mathbf{x}, P)$ with respect to P .
- For an extension to ε -optimal solutions, including certain consistency-type results see [46].

2.2.2 Quantitative stability results

The success and applicability of the quantitative stability results depend essentially on an appropriate choice of the *probability metric* d used to measure the perturbations in the probability distribution P . The probability metrics should be closely tailored to the structure of the considered stochastic program and/or to the particular type of approximation of probability distribution P . The desired results are, e.g., a Lipschitz property of the optimal value

$$d(P, P') < \eta \Rightarrow |\varphi(P) - \varphi(P')| < K\eta$$

and possibly also a Lipschitz property of the Hausdorff distance of the corresponding solution sets with respect to perturbations of P measured by d ; naturally, the Lipschitz constants depend on the chosen metric d . Again, special assumptions are needed

to get such results for optimal solutions whereas for the *sets of ε -optimal solutions*

$$\mathcal{X}_\varepsilon^*(P) := \varepsilon\text{-arg min}_{\mathbf{x} \in \mathcal{X}} F(\mathbf{x}, P) = \{\mathbf{x} \in \mathcal{X} | F(\mathbf{x}, P) \leq \varphi(P) + \varepsilon\}$$

quantitative stability results, such as

$$d_H(\mathcal{X}_\varepsilon^*(P), \mathcal{X}_\varepsilon^*(P')) \leq \frac{L}{\varepsilon} d(P, P') \quad (14)$$

for the Hausdorff distance d_H hold true under more general circumstances whenever $d(P, P')$ is small enough [2].

In the context of (1), an important class of probability metrics is based on the Kantorovich - Rubinstein functional with a continuous distance function $c : R^s \times R^s \rightarrow R_+^1$, see Chapter 5 in [43]:

$$d_c(P, Q) = \inf \left\{ \int_{R^s \times R^s} c(\mathbf{u}, \mathbf{w}) \tau(d\mathbf{u}, d\mathbf{w}) \right\} \quad (15)$$

over all finite Borel measures τ on $\Omega \times \Omega$ such that $\tau(B \times \Omega) - \tau(\Omega \times B) = P(B) - Q(B) \forall B \in \mathcal{B}$. An example is the *Fortet-Mourier metric* obtained for

$$c(\mathbf{u}, \mathbf{w}) = \|\mathbf{u} - \mathbf{w}\| \max\{1, \|\mathbf{w}\|^{p-1}, \|\mathbf{u}\|^{p-1}\}. \quad (16)$$

The distance function c has to be chosen so that the integrands $f_0(\mathbf{x}, \bullet)$ exhibit a generalized Lipschitz property with respect to ω :

$$|f_0(\mathbf{x}, \omega) - f_0(\mathbf{x}, \tilde{\omega})| \leq c(\omega, \tilde{\omega}) \forall \omega, \tilde{\omega} \in \Omega. \quad (17)$$

Then general results of parametric programming imply that small changes in the probability distribution measured by d_c result in small changes of the optimal values and of the sets of ε -optimal solutions, cf. [19] for selected theoretical results and a sketch of their application to the bond portfolio management problem. For *scenario-based* programs they can be used to quantify the desirable robustness properties also in rather complicated instances of stochastic programs with random recourse. Moreover, for two discrete probability distributions, say $P = \sum_{i=1}^I p_i \delta_{\omega^i}$, $Q = \sum_{j=1}^J q_j \delta_{\tilde{\omega}^j}$, $d_c(P, Q)$ is the optimal value of the finite-dimensional *transportation problem*

$$\min \left\{ \sum_i \sum_j c(\omega^i, \tilde{\omega}^j) \eta_{ij} \mid \sum_j \eta_{ij} = p_i \forall i, \sum_i \eta_{ij} = q_j \forall j, \eta_{ij} \geq 0 \forall i, j \right\}.$$

Application 2. A rule for deleting one scenario.

Let $P = \sum_{i=1}^I p_i \delta_{\omega^i}$ be a probability distribution which is considered to be a sufficiently precise approximation of the true one. To delete one of scenarios ω^i , $i = 1, \dots, I$, means to construct probability distribution Q_i carried by the remaining $I-1$ scenarios. This should be done so that the output of the problem based on the reduced set of

scenarios is as close as possible to the initial one. Given an already selected scenario, say, ω^l to be removed, the distance $d_c(P, Q_l)$ depends on the known probabilities $p_i \forall i$, on known distances $c_{ij} := c(\omega^i, \omega^j) \forall i, j$ and on probabilities $q_j, j \neq l, j = 1, \dots, I$, to be fixed in an optimal way. The dual to the transportation problem can be solved explicitly and its optimal value $d_c(P, Q_l) = p_l \min_{j \neq l} c_{lj}$. The *corresponding deletion rule* is then:

"Remove scenario ω^k with $k \in \arg \min_{l=1, \dots, I} [p_l \min_{j \neq l} c_{lj}]$ and redistribute its probability p_k to that scenario ω^l for which the distance $c_{kl} = \min_{j \neq k} c_{kj}$."

See [26] for an application of this rule and [18] for other related results.

Comments

- Extensions of these quantitative stability results to multistage stochastic programs and to $F(\mathbf{x}, P)$ nonlinear in P are in progress.
- Precision of the bounds for the difference of the optimal values and for the distance of the sets of ε -optimal solutions depends on possibility to get a sufficiently tight estimate of the Lipschitz (or Hölder) constants, such as L in (14). Nevertheless, we have seen that in spite of this difficulty, quantitative stability results can help in designing a discrete approximation P' of P which is representative enough in the sense that the obtained solutions enjoy plausible robustness properties.
- The Lipschitz property (17) cannot be expected for discontinuous $f_0(\mathbf{x}, \bullet)$; another type of distance, e.g., the *variational distance (discrepancy)* fits better to integer stochastic programs or to problems with probability constraints; for a discussion see [52].

2.3 Contamination technique

Again, it is expedient to model the perturbations of the probability distribution P using a finite dimensional parameter. This can be done by the contamination method.

Contamination means to model the perturbed probability distribution as

$$P_\lambda = (1 - \lambda)P + \lambda Q, \quad 0 \leq \lambda \leq 1, \quad (18)$$

the *probability distribution P contaminated by the probability distribution Q* . The contamination neighborhood

$$\mathcal{O}_\lambda(P) := \{\hat{P} | \hat{P} = (1 - \lambda)P + \lambda Q, \forall \text{ probability distributions } Q\}$$

is not a neighborhood in the topological sense, but for λ small enough, the contaminated distributions fall into a neighborhood of P . For fixed probability distributions P, Q , we denote $F(\mathbf{x}, P_\lambda)$ the objective function in (1) computed for the contaminated distribution and by

$$\varphi(\lambda) := \min_{\mathbf{x} \in \mathcal{X}} F(\mathbf{x}, P_\lambda)$$

the optimal value function. Notice that for the expectation-type of stochastic programs, P_λ is *linear* in λ so that $F(\mathbf{x}, P_\lambda)$ is linear in the scalar parameter λ , $\frac{d}{d\lambda}F(\mathbf{x}, P_\lambda) = F(\mathbf{x}, Q) - F(\mathbf{x}, P)$ and classical approaches of parametric programming can be exploited. For instance, by application of [25] in case of convex expectation-type stochastic programs (1), with $\mathcal{X}^*(P) \neq \emptyset$ and compact, $\mathcal{X}^*(Q) \neq \emptyset$, the optimal value $\varphi(\lambda)$ is a finite concave function on $[0, 1]$, continuous at 0, whose derivative at $\lambda = 0^+$

$$\varphi'(0^+) = \min_{\mathbf{x} \in \mathcal{X}^*(P)} F(\mathbf{x}, Q) - \varphi(0). \quad (19)$$

Bounds on the optimal value $\varphi(\lambda)$ for an arbitrary $\lambda \in [0, 1]$ then follow by properties of concave functions:

$$(1 - \lambda)\varphi(0) + \lambda\varphi(1) \leq \varphi(\lambda) \leq \varphi(0) + \lambda\varphi'(0^+) \quad \forall \lambda \in [0, 1]. \quad (20)$$

An upper bound for the derivative (19) equals $F(\mathbf{x}(P), Q) - \varphi(0)$ where $\mathbf{x}(P)$ is an *arbitrary* optimal solution of the initial problem (1) obtained for the probability distribution P ; if the optimal solution is unique, this upper bound is attained. Hence, evaluation of bounds in (20) requires solution of another stochastic program of the type (1) for the new distribution Q to get $\varphi(1)$ and evaluation of the expectation $F(\mathbf{x}(P), Q)$ at an already known optimal solution $\mathbf{x}(P)$ of the initial problem (1) but for the contaminating distribution Q .

This result can be generalized to $F(\mathbf{x}, P)$ convex in \mathbf{x} and *concave* in P , see [14]: Under the assumptions listed above, $\varphi(\lambda)$ remains concave on $[0, 1]$, continuous at 0. Additional assumptions are needed to get the existence of the derivative

$$\varphi'(0^+) = \min_{\mathbf{x} \in \mathcal{X}^*(P)} \frac{d}{d\lambda} F(\mathbf{x}, \lambda)|_{\lambda=0^+}.$$

Possible generalizations to *nonconvex* problems (1) can be based on the corresponding qualitative stability result quoted in Section 2.2.1; see Theorem 8.2 of [10].

Application 3. Additional scenario for SLP with integer recourse.

To simplify the notation we shall present the result for SLP with complete integer recourse and random right-hand sides; see [7] for generalizations and details. It means that the random objective function in (1) is

$$f_0(\mathbf{x}, \omega) = \mathbf{c}^\top \mathbf{x} + \Phi(\omega - \mathbf{A}\mathbf{x})$$

with

$$\Phi(\mathbf{z}) = \min \left\{ \mathbf{q}^\top \mathbf{y} + \mathbf{q}'^\top \mathbf{y}' : \mathbf{W}\mathbf{y} + \mathbf{W}'\mathbf{y}' = \mathbf{z}, \mathbf{y}' \geq \mathbf{0}, \mathbf{y} \geq \mathbf{0}, \mathbf{y} \text{ integer} \right\}$$

where \mathbf{q}, \mathbf{q}' and the matrices \mathbf{W} and \mathbf{W}' have rational entries.

Under common assumptions of integrability of $\|\omega\|$, of complete recourse and for $\{\mathbf{u} | \mathbf{W}^\top \mathbf{u} \leq \mathbf{q}, \mathbf{W}'^\top \mathbf{u} \leq \mathbf{q}'\} \neq \emptyset$, it is possible to prove that $F(\bullet, P)$ is a real-valued lower semicontinuous function on R^n ; see [51], Proposition 3.1. For absolutely continuous P and for probability distributions belonging to the class $\Delta_{p,K}$ on R^m , the general stability theorem holds true ([51], Proposition 4.1). Then the Gâteaux derivative $\varphi'_G(0^+)$ of φ_G at P in the direction of $Q - P$ exists for an arbitrary probability distribution $Q \in \Delta_{p,K}(R^m)$ and is given by

$$\varphi'_G(0^+) = \varphi_G(P; Q - P) = \min_{\mathbf{x} \in \mathcal{X}_G^*(P)} F(\mathbf{x}, Q) - \varphi_G(P).$$

Hence, the bounds (20) apply for $\varphi_G(\lambda)$. Notice that the result holds true also for Q a degenerated or discrete probability distribution.

Comments

- The contamination method does not require any specific properties of the probability distribution P , and it does not require any convexity-type assumptions about $f_0(\mathbf{x}, \bullet)$. This means that it can be applied also to *random recourse* and to *multistage* stochastic programs, cf. [13].

- It is applicable for objectives which are not linear in P , cf. [14, 15].

- It can be relaxed to *nonconvex optimization problems*. This opens a possibility to apply it in the context of stochastic programs with *integer* recourse.

- The general quantitative stability results apply also to the sets of ε -optimal solutions of contaminated stochastic programs. For instance, using the Fortet-Mourier metric with the distance function (16), we get distance $d_c(P, P_\lambda) = \lambda d_c(P, Q)$ to be used in (14) for λ small enough. For a quantitative stability result concerning sets of optimal solutions of contaminated two-stage stochastic programs under special assumptions see Proposition 2.10 of [49].

In the framework of scenario-based stochastic programs, contamination technique can be used to study the influence of the assigned values of probabilities p_s and of the whole probability structure, including additional stages and additional scenarios or branches of the scenario tree on the optimal value; see [13, 14]. For stability studies with respect to small changes in the underlying probability distribution, small values of the contamination parameter λ are typical. The choice of λ may reflect the degree of confidence in expert opinions represented as the contaminating distribution Q , and so on. Using a contaminating distribution Q carried by additional scenarios or branches of the scenario tree, one can study the influence of including these additional "out-of-sample" scenarios; cf. [17] for application in portfolio optimization and [15] for an application to robust optimization models. By a suitable choice of Q , criteria on a right number of stages can be tested, the response on an increasing importance of a scenario can be quantified, etc. See [16] for examples and discussions.

2.4 Bounds and worst case analysis

The purpose of this Section is not to list various bounding techniques which aim at designing approximation schemes and stopping rules for numerical procedures, e.g., [4, 29]. We will only complement the already presented results related to the output analysis with respect to the input probability distribution.

The worst case analysis is an additional approach which is useful in case of an *incomplete information about the probability distribution* P and also for other than empirical approximations of P . One assumes that P belongs to a family \mathcal{P} of probability distributions identified, e.g., by known values of some moments, by a given support, by qualitative properties, such as unimodality [9], or by unprecise values of probabilities of expert scenarios [5], etc.

In the context of (1) with $F(\mathbf{x}, P) := E_P f_0(\mathbf{x}, \omega)$, one can try to construct *minmin* and *minmax* bounds

$$\min_{\mathbf{x} \in \mathcal{X}} \inf_{P \in \mathcal{P}} F(\mathbf{x}, P) \leq \varphi(P) \leq \min_{\mathbf{x} \in \mathcal{X}} \sup_{P \in \mathcal{P}} F(\mathbf{x}, P) \quad \forall P \in \mathcal{P} \quad (21)$$

on the optimal value of the true program to get information about robustness of the optimal value within the considered family of probability distributions.

The objective functions of the inner minimization and maximization problems

$$\inf_{P \in \mathcal{P}} E_P f_0(\mathbf{x}, \omega) \text{ and } \sup_{P \in \mathcal{P}} E_P f_0(\mathbf{x}, \omega)$$

are linear in P , which means that for convex, compact set \mathcal{P} , the infimum and supremum are attained and the optimal *best case* and *worst case* probability distributions P^* , $P^{**} \in \mathcal{P}$ are extremal points of \mathcal{P} .

In the framework of the moment problem, see [33], these extremal points are well described for \mathcal{P} defined by a given support and by known values of certain generalized moments: For admissible moment values, the extremal distributions are discrete ones, concentrated in a modest number of points; hence, the bounds (21) follow by solution of a *scenario-based program*. However, extremal distributions *independent* of the decisions \mathbf{x} appear only exceptionally, under special assumptions (e.g., convexity, concavity or saddle property) about the integrand $f_0(\mathbf{x}, \bullet)$ and about the families of distribution functions, e.g., for those with given support and expectations: Recall the Jensen inequality valid for all probability distributions P with a prescribed expected value $E_P \omega = \bar{\omega} \quad \forall P \in \mathcal{P}$ and for convex $f_0(\mathbf{x}, \bullet)$, so that the best case distribution, independently of \mathbf{x} , is degenerated, concentrated at $\bar{\omega}$. On the other hand, for convex $f_0(\mathbf{x}, \bullet)$ and compact convex polyhedral support Ω , the worst case distributions are concentrated at the extremal points of Ω . This is an old result of Edmundson and Madansky [38] which provides the worst case distribution independent of \mathbf{x} under special assumptions, e.g., when Ω is a simplex or a Cartesian product of simplices. In general case, one has to apply a numerical method for solving the minmin and minmax problems in (21); see [24].

A related, though less ambitious problem is to get bounds on the performance of an optimal solution $\mathbf{x}(P)$ obtained for a probability distribution $P \in \mathcal{P}$ using the corresponding worst case and best case probability distributions from \mathcal{P} . This leads to bounds which are then exploited in various computational schemes, e.g. [30] for two-stage stochastic programs. Also here a tractable procedure for the (repeated) evaluation of bounds requires certain convexity properties of the function $f_0(\mathbf{x}, \omega)$ with respect to ω and a special type of family \mathcal{P} .

The required convexity-type property of $f_0(\mathbf{x}, \bullet)$ restricts thus the applicability of this approach to fixed recourse stochastic programs whose second-stage coefficients are supposed to depend on the random parameters in a *linear* way. An extension to software development for multiperiod and multistage stochastic programs can be found e.g. in [23]. For multistage stochastic programs, convexity or saddle property of $f_0(\mathbf{x}, \bullet)$ depends upon special additional assumptions about the probability distribution of the stochastic data process: It can be obtained for fixed recourse and interstage independence. In presence of interstage dependences, even with randomness entering only the right-hand sides, the convexity property is known to be valid only for special probability distributions of the random right-hand sides, namely, for the multinormal ones.

Comments

- The worst case analysis is optimal value oriented and has been developed for expectation type objective functions, i.e., for $F(\mathbf{x}, P)$ linear in P . Nevertheless, the general methodology related with the moment problem can be applied also to more general types of F .
- For reasons of numerical tractability, however, it is limited to random objective functions $f_0(\mathbf{x}, \bullet)$ which fulfil certain convexity-type properties with respect to ω and to special families of probability distributions.

3 Concluding remarks

For purposes of output analysis, consistency and qualitative stability provide necessary background properties of the solved approximated stochastic programs and they are needed for development of quantitative results. In general, it is much easier to obtain such results for the optimal value than for optimal solutions and it makes sense to bring forward approaches to analysis of ε -optimal solutions.

The asymptotic results of Section 2.1 for (wide-sense) empirical probability distributions allow to construct approximate tests and confidence intervals for the true optimal value. In general, however, this is not the case of optimal solutions. For the empirical stochastic programs, an open possibility is to apply the normalized convergence results of [21], the asymptotic dominance property by [41] or to elaborate further the already existing large deviations results [31, 55].

Also quantitative stability results, see Subsection 2.2.2 offer a certain type of

bounds for optimal value and sets of ε -optimal solutions, depending on the probability metrics and numerical values of the Lipschitz or Hölder constants.

Contamination, see Section 2.3, provides easily computable bounds for optimal value under rather weak assumptions about the random objective f_0 . It has not yet been put to the test for quantitative stability of the sets of ε -optimal solutions of contaminated problems, still it is possible to rely on general quantitative stability results.

The worst case analysis is optimal value oriented and the directly applicable results require certain convexity-type properties of the integrand $f_0(\mathbf{x}, \omega)$ with respect to ω .

For most of the presented approaches, convexity of the objective function $F(\mathbf{x}, P)$ with respect to the decision variables \mathbf{x} and its linearity with respect to the probability distribution P do not mean any serious limitation, just a simplification.

Acknowledgement. This work was partly supported by research project "Mathematical Methods in Stochastics" - MSM113200008 and by grants 201/99/0264 and 402/99/1136 of the Grant Agency of the Czech Republic.

References

- [1] Artstein, Z. and Wets, R. J.-B. (1994), "Stability results for stochastic programs and sensors, allowing for discontinuous objective functions," *SIAM J. Optimization*, **4**, 537–550.
- [2] Attouch, H. and Wets, R. J.-B. (1993), "Quantitative stability of variational systems: III. ϵ -approximate solutions," *Math. Progr.*, **61**, 197–214.
- [3] Bank, B., Guddat, J., Klatte, D., Kummer, B. and Tammer, K. (1988), *Non-Linear Parametric Optimization*, Akademie-Verlag, Berlin.
- [4] Birge, J. R. and Wets, R. J.-B. (1986), "Designing approximation schemes for stochastic optimization problems, in particular for stochastic programs with recourse," *Math. Progr. Study*, **27**, 54–102.
- [5] Bühler, W. (1981), "Flexible Investitions- und Finanzplanung bei unvollkommen bekannten Übergangswahrscheinlichkeiten," *OR Spektrum*, **2**, 207–221.
- [6] Dempster, M. A. H. (ed.) (1980), *Stochastic Programming*, Academic Press, London.
- [7] Dobiáš, P. (2000), Manuscript of PhD thesis, Charles University, Prague.
- [8] Dupačová, J. (1984), "Stability in stochastic programming with recourse – estimated parameters," *Math. Progr.*, **28**, 72–83.

- [9] Dupačová, J. (1987), “The minimax approach to stochastic programming and an illustrative application,” *Stochastics*, **20**, 73–88.
- [10] Dupačová, J. (1990), “Stability and sensitivity analysis for stochastic programming,” *Annals of Oper. Res.*, **27**, 115–142.
- [11] Dupačová, J. (1991), “On statistical sensitivity analysis in stochastic programming,” *Annals of Oper. Res.*, **30**, 199–214.
- [12] Dupačová, J. (1991), “On non-normal asymptotic behavior of optimal solutions for stochastic programming problems and on related problems of mathematical statistics,” *Kybernetika*, **27**, 38–52.
- [13] Dupačová, J. (1995), “Postoptimality for multistage stochastic linear programs,” *Annals of Oper. Res.*, **56**, 65–78.
- [14] Dupačová, J. (1996), “Scenario based stochastic programs: Resistance with respect to sample,” *Annals of Oper. Res.*, **64**, 21–38.
- [15] Dupačová, J. (1998), “Reflections on robust optimization,” in: *Stochastic Programming Methods and Technical Applications* (K. Marti and P. Kall, eds.), LNEMS **458**, Springer, Berlin, pp. 111–127.
- [16] Dupačová, J. (1999), “Portfolio optimization via stochastic programming: Methods of output analysis,” *MMOR*, **50**, 245–270.
- [17] Dupačová, J., Bertocchi, M. and Moriggia, V. (1998), “Postoptimality for scenario based financial models with an application to bond portfolio management,” in: *World Wide Asset and Liability Modeling* (W. Ziemba and J. Mulvey, eds.), Cambridge Univ. Press, pp. 263–285.
- [18] Dupačová, J., Gröwe-Kuska, N. and Römisch, W. (2000), “Scenario reduction in stochastic programming: An approach using probability metrics,” Manuscript, Humboldt University, Berlin.
- [19] Dupačová, J. and Römisch, W. (1998), “Quantitative stability for scenario-based stochastic programs,” in: *Prague Stochastics’98* (M. Hušková et al., eds.), JČMF, Prague, pp. 119–124.
- [20] Dupačová, J. and Wets, R. J.-B. (1988), “Asymptotic behavior of statistical estimators and of optimal solutions of stochastic optimization problems,” *Ann. Statist.*, **16**, 1517–1549.
- [21] Ermoliev, Yu. and Norkin, V. I. (1991), “Normalized convergence in stochastic optimization,” *Annals of Oper. Res.*, **30**, 187–198.

- [22] Ermoliev, Yu. and Wets, R. J.-B., eds. (1988), *Numerical Techniques for Stochastic Optimization Problems*, Springer, Berlin.
- [23] Frauendorfer, K. (1996), “Barycentric scenario trees in convex multistage stochastic programming,” *Math. Progr.*, **75**, 277–293.
- [24] Gaivoronski, A. A. (1991), “A numerical method for solving stochastic programming problems with moment constraints on a distribution function,” *Annals of Oper. Res.*, **31**, 347–369.
- [25] Gol’shtein, E. G. (1970), *Vypukloje Programirovanie. Elementy Teoriji*, Nauka, Moscow. [*Theory of Convex Programming*, Translations of Mathematical Monographs, **36**, American Mathematical Society, Providence RI, 1972].
- [26] Gröwe-Kuska, N., Kiwiel, K. C., Nowak, M. P., Römisch, W. and Wegner, I. (1999), “Power management in a hydro-thermal system under uncertainty by Lagrangian relaxation,” Preprint 99–19, Institut für Mathematik, Humboldt University, Berlin, available at *Stochastic Programming E-Print Series* <http://dochostr.rz.hu-berlin.de/speps/>.
- [27] Kall, P. (1976), *Stochastic Linear Programming*, Springer, Berlin.
- [28] Kall, P. (1987), “On approximations and stability in stochastic programming,” in: *Parametric Optimization and Related Topics* (J. Guddat et al., eds.), Math. Research Band **35**, Akademie-Verlag, Berlin, pp. 387–347.
- [29] Kall, P., Frauendorfer, K. and A Ruszczyński (1988), “Approximation techniques in stochastic programming,” Chapter 2 in [22].
- [30] Kall, P. and Stoyan, D. (1982), “Solving stochastic programming problems with recourse including error bounds,” *Math. Operationsforschung Statist. Ser. Optimization*, **13**, 431–447.
- [31] Kaniovski, Y. M., King, A. J. and Wets, R. J.-B. (1995), “Probabilistic bounds (via large deviations) for the solutions of stochastic programming problems,” *Annals of Oper. Res.*, **56**, 189–208.
- [32] Kaňková, V. (1977), “Optimum solution of a stochastic optimization problem with unknown parameters,” in: *Transactions of the 7th Prague Conference 1974*, Academia, Prague, pp.239–244.
- [33] Kemperman, J. M. B. (1968), “The general moment problem, a geometric approach,” *Ann. Math. Statist.*, **39**, 93–122.
- [34] King, A. J. (1989), “Generalized delta theorems for multivalued mappings and measurable selections,” *Math. of Oper. Res.*, **14**, 720–736.

- [35] King, A. J. and Rockafellar, R. T. (1993), "Asymptotic theory for solutions in statistical estimation and stochastic programming," *Math. of Oper. Res.*, **18**, 148–162.
- [36] King, A. J. and Wets, R. J.-B. (1990), "Epi-consistency of convex stochastic programs," *Stochastics and Stochastics Reports*, **34**, 83–92.
- [37] Korf, L. A. and Wets, R. J.-B. (1999), "An ergodic theorem for stochastic programming problems," Preprint, University of California, Davis.
- [38] Madansky, A. (1959), "Bounds on the expectation of a convex function of a multivariate random variable," *Ann. Math. Statist.*, **30**, 743–746.
- [39] Lucchetti, R. and Wets, R. J.-B. (1993), "Convergence of minima of integral functionals, with applications to optimal control and stochastic optimization," *Statistics & Decisions*, **11**, 69–84.
- [40] Mulvey, J. M., Vanderbei, R. J. and Zenios, S. A. (1995), "Robust optimization of large scale systems," *Oper. Res.*, **43**, 264–281.
- [41] Pflug, G. Ch. (1992), "Asymptotic dominance and confidence for solutions of stochastic programs," *CZ Journal for Oper. Res.*, **1**, 21–30.
- [42] Pflug, G. Ch. (1995), "Asymptotic stochastic programs," *Math. of Oper. Res.*, **4**, 769–789.
- [43] Rachev, S. T. (1991), *Probability Metrics and the Stability of Stochastic Models*, Wiley, New York.
- [44] Robinson, S. M. (1987), "Local structure of feasible sets in nonlinear programming. Part III: Stability and sensitivity," *Math. Progr. Study*, **30**, 45–66.
- [45] Robinson, S. M. (1987), "Local epi-continuity and local optimization," *Math. Progr.*, **37**, 208–222.
- [46] Robinson, S. M. (1996), "Analysis of sample-path optimization," *Math. Oper. Res.*, **21**, 513–528.
- [47] Robinson, S. M. and Wets, R. J.-B. (1987), "Stability in two-stage stochastic programming," *SIAM J. Control and Optimization*, **25**, 1409–1416.
- [48] Rockafellar, R. T. and Wets, R. J.-B. (1997), *Variational Analysis*, Springer, New York.
- [49] Römisch, W. and Schultz, R. (1996), "Lipschitz stability for stochastic programs with complete recourse," *SIAM J. Optimization*, **6**, 531–547.

- [50] Rubinstein, R. Y. and Shapiro, A. (1993), *Discrete Event Systems: Sensitivity Analysis and Stochastic Optimization by the Score Function Method*, Wiley, New York.
- [51] Schultz, R. (1995), "On structure and stability in stochastic programs with random technology matrix and complete integer recourse," *Math. Progr.*, **70**, 73–89.
- [52] Schultz, R. (1996), "Rates of convergence in stochastic programs with complete integer recourse," *SIAM J. Optimization*, **6**, 1138–1152.
- [53] Serfling, R. J. (1980), *Approximation Theorems in Mathematical Statistics*, Wiley, New York.
- [54] Shapiro, A. (1991), "Asymptotic analysis of stochastic programs," *Annals of Oper. Res.*, **30**, 169–186.
- [55] Shapiro, A. and Homem-de-Mello, T. (1999), "On rate of convergence of optimal solutions of Monte Carlo approximations of stochastic programs," Preprint available at *Stochastic Programming E-Print Series* <http://dochoost.rz.huberlin.de/speps/>.
- [56] Tlusty, P. (1991), "Deviations from asymptotic normality for optimal solutions of stochastic programming problems" [in Czech], PhD thesis, Charles University, Prague.
- [57] Wang, L. and Wang, J. (1999), "Limit distribution of statistical estimators for stochastic programs with dependent samples," *ZAMM*, **19**, 257–266.
- [58] Wets, R. J.-B. (1979), "A statistical approach to the solution of stochastic programs with (convex) simple recourse," Working Paper, Univ. of Kentucky.
- [59] Wets, R. J.-B. (1983), "Stochastic programming: Solution techniques and approximation schemes," in: *Mathematical Programming – The State of the Art Bonn 1982* (J. Bachem et al., eds.), Springer, Berlin, pp. 566–603.
- [60] Wets, R. J.-B. (1989), "Stochastic programming," Chapter VII in: *Handbooks in OR & MS 1* (G. L. Nemhauser et al., eds.), Elsevier, pp. 573–629.
- [61] Wets, R. J.-B. (1996), "Challenges in stochastic programming," *Math. Progr.*, **75**, 115–135.